nature microbiology

LETTERS

https://doi.org/10.1038/s41564-019-0502-x

Silicon limitation facilitates virus infection and mortality of marine diatoms

Chana F. Kranzler[1], Jeffrey W. Krause[2,3], Mark A. Brzezinski[4], Bethanie R. Edwards[5], William P. Biggs[1], Michael Maniscalco[4], John P. McCrow[6], Benjamin A. S. Van Mooy[7], Kay D. Bidle[1], Andrew E. Allen[6,8] and Kimberlee Thamatrakoln[1*]

Diatoms are among the most globally distributed and ecologically successful organisms in the modern ocean, contributing upwards of 40% of total marine primary productivity[1,2]. By converting dissolved silicon into biogenic silica, and photosynthetically fixing carbon dioxide into particulate organic carbon, diatoms effectively couple the silicon (Si) and carbon cycles and ballast substantial vertical flux of carbon out of the euphotic zone into the mesopelagic and deep ocean[3–5]. Viruses are key players in ocean biogeochemical cycles[6,7], yet little is known about how viral infection specifically impacts diatom populations. Here, we show that Si limitation facilitates virus infection and mortality in diatoms in the highly productive coastal waters of the California Current Ecosystem. Using metatranscriptomic analysis of cell-associated diatom viruses and targeted quantification of extracellular viruses, we found a link between Si stress and the early, active and lytic stages of viral infection. This relationship was also observed in cultures of the bloom-forming diatom *Chaetoceros tenuissimus*, where Si stress accelerated virus-induced mortality. Together, these findings contextualize viruses within the ecophysiological framework of Si availability and diatom-mediated biogeochemical cycling.

Silicon (Si) cycling and diatom production are inextricably intertwined. The availability of dissolved silicon (dSi) sets the upper limit on diatom productivity and governs the distribution of diatoms in the global ocean[8]. Given that as much as 80–90% of biogenic silica ($bSiO_2$) in the euphotic zone can be occluded within diatom detritus[9] and that up to 70% of new diatom production is supported through the dissolution of this detrital silica[10], the processes that contribute to diatom mortality and mediate $bSiO_2$ remineralization are essential components of upper-ocean Si and carbon (C) cycling.

Marine viruses are estimated to turn over more than a quarter of photosynthetically fixed carbon, fuelling the microbial loop and circumventing the transfer of organic carbon to higher trophic levels and the deep ocean[11]. However, recent work suggests that, counter to this 'viral-shunt hypothesis', viruses may also stimulate aggregation, sinking and export[12]. Despite the importance of viral infection to phytoplankton mortality[13], the environmental factors that impact these dynamics and the biogeochemical consequences of these interactions are not well defined. Diatom viruses have the potential to substantially impact both the Si and C cycles; however, virus-induced losses in diatom populations are poorly understood. With virions approximately 20–40 nm in diameter and genomes comprised of either single-stranded (ss) RNA or ssDNA[14], diatom viruses are among the smallest on the planet and distinct from the better characterized and more readily detected double-stranded (ds) DNA algal viruses (for example *Phycodnaviridae*) and bacteriophages. However, recent estimates suggest RNA viruses comprise a substantial fraction of the marine virus community and may be as impactful as DNA viruses in the turnover of organic matter[15]. Advances in metagenomics and metatranscriptomics have facilitated the detection of marine ssDNA- and ssRNA-containing viruses, thus enabling the interrogation of diatom host–virus interactions in natural populations[16,17].

We explored viral infection during the early summer diatom blooms in the California Current Ecosystem as part of the 'DYE labelling of diatom silica' (DYEatom) cruise (research vessel (RV) *Point Sur*, PS1312). This region is characterized by seasonal, wind-driven coastal upwelling of nutrient-rich waters during the spring and summer seasons that sustains extensive diatom blooms[18]. Phytoplankton populations were sampled at multiple coastal and offshore stations near Monterey Bay and Point Reyes during June–July 2013 (Fig. 1). Satellite imagery of chlorophyll depicts high phytoplankton biomass along the coast during the sampling period (Fig. 1a), with in situ chlorophyll a concentrations reaching 14 μg l^{-1} (Supplementary Table 1). Compared with the offshore stations Station 6 (St6) and St9, the coastal stations were characterized by high $bSiO_2$ (4.6–21.7 μmol l^{-1} Si), high diatom transcript abundance (Supplementary Data 1) and diatom-dominated 18S ribosomal RNA (84–97%; Fig. 1b), indicating high diatom abundance at these stations. In Monterey Bay, *Pseudo-nitzschia* was the dominant genus, whereas nearshore stations close to Point Reyes were more diverse, with *Chaetoceros*, *Thalassiosira* and *Pseudo-nitzschia* genera present (Fig. 1c). The two offshore stations (St6 and St9) seemed to be dominated by dinoflagellates and haptophytes, as indicated by low $bSiO_2$ levels (<1.3 μmol l^{-1} Si) and a high proportion of these groups in the 18S rRNA amplicons (Fig. 1b). Kinetic limitation of Si uptake was quantified as a proxy for Si stress (see Methods), with values approaching zero indicative of severe Si limitation[19]. In Monterey Bay, Si stress was variable, ranging from none (V_{amb}:V_{enh} ≈ 1) to moderate (V_{amb}:V_{enh} = 0.28–0.77), whereas diatom-dominated stations were characterized by severe Si stress (V_{amb}:V_{enh} < 0.1; Fig. 1d) near Point Reyes. Phaeophytin, a degradation product of chlorophyll and an indicator of mortality,

[1]Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA. [2]Dauphin Island Sea Lab, Dauphin Island, AL, USA. [3]Department of Marine Sciences, University of South Alabama, Mobile, AL, USA. [4]Marine Science Institute, University of California, Santa Barbara, CA, USA. [5]Department of Earth and Planetary Science, University of California, Berkeley, CA, USA. [6]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA, USA. [7]Woods Hole Oceanographic Institution, Woods Hole, MA, USA. [8]Scripps Institution of Oceanography, University of California, San Diego, CA, USA. *e-mail: thamat@marine.rutgers.edu

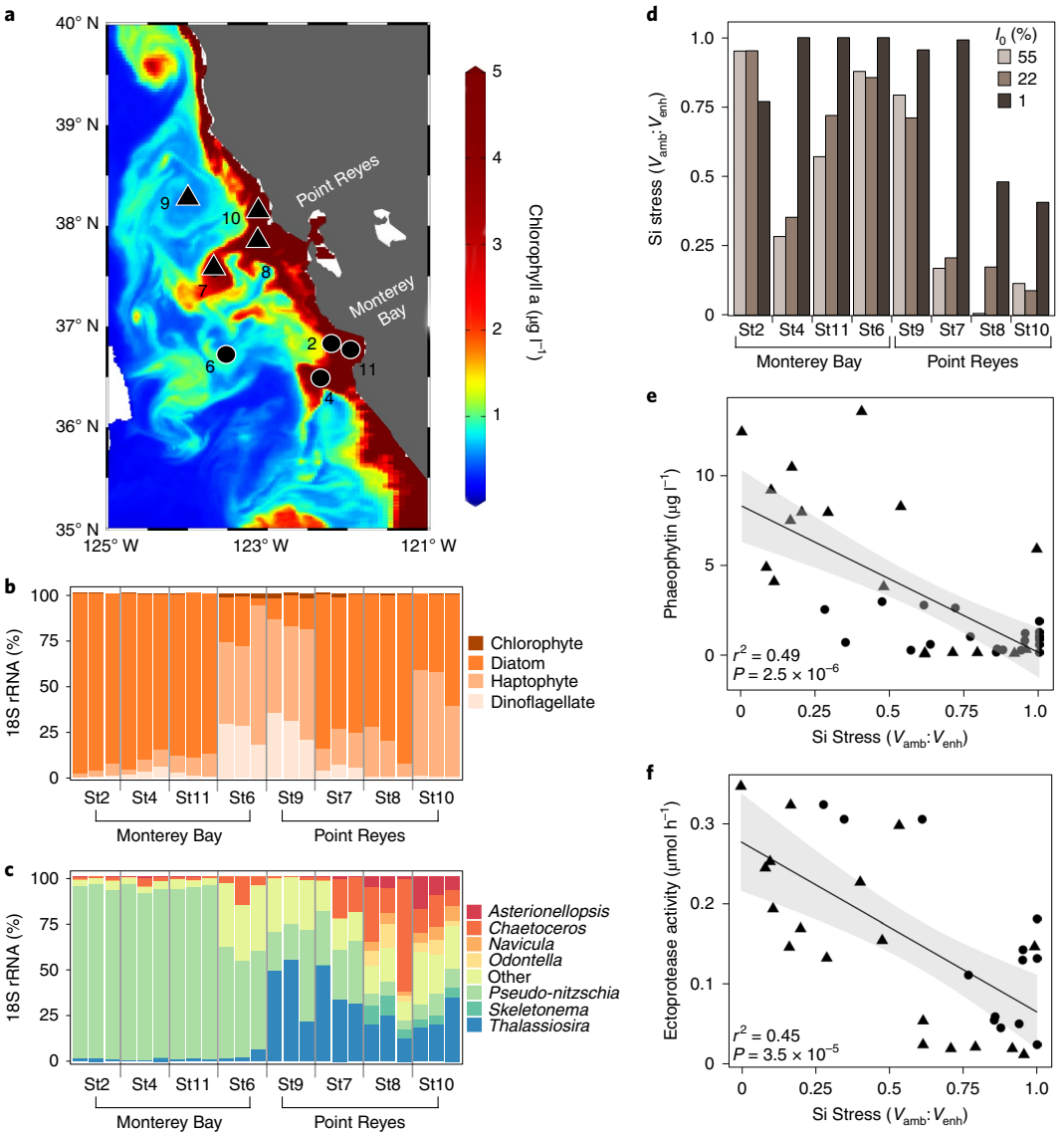


Fig. 1 | Station map, phytoplankton communities, Si stress and indicators of mortality during the DYEatom cruise in the California Current Ecosystem. **a**, Station locations overlaid on Level-2 MODIS surface chlorophyll concentrations for 29 June 2013 (oceancolor.gsfc.nasa.gov). **b**,**c**, Relative 18S rRNA abundance of the major eukaryotic phytoplankton groups (**b**) and diatom genera (**c**) at each station. Data from three depths, corresponding to 55, 22 and 1% surface irradiance (I_0) are shown within each station from left to right. **d**, Si stress (V_{amb}:V_{enh}) at each station, defined as the ratio between specific uptake rates at ambient and saturating $Si(OH)_4$ (Supplementary Table 1), with values approaching zero indicative of severe Si stress. **e**,**f**, Relationship between Si stress and phaeophytin (**e**), and ectoproteolytic activity (**f**) determined using leucine-AMC ($n = 35$ and 31 independent samples, respectively). The lines of the best fit with 95% confidence intervals (grey shading) are shown describing linear regressions with the r^2 and P values (analysis of variance). The circles and triangles represent the Monterey Bay and Point Reyes stations, respectively.

correlated significantly with Si stress (Fig. 1e). This correlation was not observed between chlorophyll a and Si stress (coefficient of multiple correlation (r^2) = 0.01, P = 0.5), suggesting that total phytoplankton biomass did not drive this relationship. Elevated ectoprotease activity, a response that has been attributed to cell death and enhanced bacterial activity during bloom demise[20,21], was also significantly correlated with Si stress (Fig. 1f). Regionally, the Point Reyes stations dominated the upper limit of these relationships (Fig. 1e,f), with distinctly higher phaeophytin and ectoprotease activity relative to chlorophyll a (Supplementary Fig. 1a–c). Together, these data suggest that the communities near Point Reyes experienced a recent mortality event that coincided with severe Si stress.

To explore the presence of diatom viruses in these communities, we used metatranscriptomic analysis of cell-associated viruses and quantitative PCR with reverse transcription (qRT–PCR) of free, extracellular viruses. Diverse ssRNA viruses were identified at all stations and depths, with the two most abundant groups being diatom viruses (*Bacillarnavirus)* and an unclassified virus thought to infect an oomycete closely related to *Phytopthera infestans* (Supplementary Fig. 2a). A homology-based query using the taxonomic gene marker RNA-dependent RNA polymerase (RdRP) identified five distinct diatom-virus-like contigs (E-value < 10^{-3}) in 20 of the 24 samples (Supplementary Fig. 2b). However, phylogenetic analysis demonstrated that only three of these five contigs grouped among known diatom viruses (Supplementary Fig. 2c) and were thus selected for quantitative analysis (see Methods). The abundance of both total cell-associated and extracellular diatom viruses were consistently high across the Monterey Bay coastal

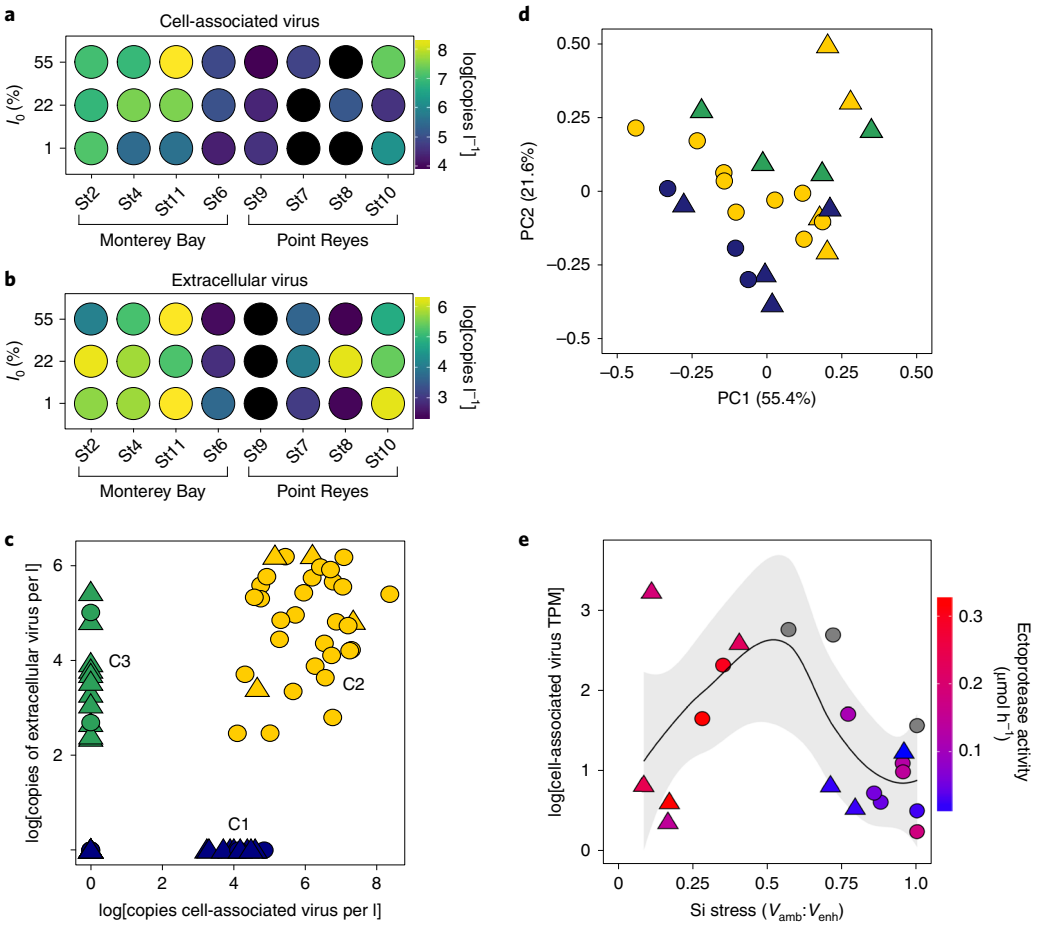


Fig. 2 | Patterns of virus infection in diatom populations in the California Current Ecosystem. a,**b**, Sum of the three cell-associated diatom virus RdRP contigs (**a**) and of these same contigs identified in the extracellular fraction (**b**) at each station at depths corresponding to 55, 22 and 1% surface irradiance (I_o). The black circles represent stations in which no diatom viruses were detected. **c**, Extracellular and cell-associated virus abundance for each viral contig. **d**, PCA biplot of the environmental metadata including nitrate/nitrite, phosphate, silicate, Si stress, chlorophyll a, $bSiO_2$, phaeophytin and percentage of surface irradiance ($n = 24$ independent samples). The colours in **c**,**d** depict the *k*-means clustering of the cell-associated and extracellular diatom virus abundance for each contig (C1, blue; C2, yellow and C3, green). **e**, Normalized cell-associated diatom-virus abundance versus Si stress in the Monterey Bay and Point Reyes stations ($n = 20$ independent samples). Line of best fit (black line) and 95% confidence intervals (grey shading) describe a LOESS regression. Colours depict ectoprotease activity determined using leucine-AMC. Grey symbols represent samples for which activity was not measured. TPM, transcripts per million. The circles and triangles in **c**–**e** represent the Monterey Bay and Point Reyes stations, respectively.

stations (Fig. 2a,b). Offshore St6 and St9 exhibited low or undetectable abundance of cell-associated and extracellular diatom viruses (Fig. 2a,b). In contrast, the abundance of cell-associated diatom viruses at the Point Reyes stations was highly variable, ranging from no detectable infection (St7 and St8) to $>10^7$ copies l^{-1} (St10; Fig. 2a). Interestingly, extracellular viruses were present at all of these Point Reyes stations, ranging from about 10^2 to 10^6 copies l^{-1} (Fig. 2b).

Analysis of the abundance of cell-associated and extracellular virus of each diatom virus yielded three distinct populations (Fig. 2c). Cluster 1 (C1) was characterized by moderate cell-associated virus abundance with no detectable extracellular viruses, suggesting an early stage of a developing infection that had not yet progressed to host lysis. C2 had high levels of both cell-associated and extracellular viruses suggesting an active, prolific infection in these populations. Populations in C3 had undetectable levels of cell-associated viruses but high levels of extracellular viruses, suggesting a post-lytic stage of infection in which most of the host population had lysed. The distribution of stations within each cluster is consistent with an active infection in Monterey Bay and a virus-mediated lytic event near Point Reyes. Principal component analysis (PCA) of the environmental metadata (Supplementary Table 1) yielded a biplot accounting for 77% of the variance (Fig. 2d), in which C1, C2 and C3 populations (early, active and lytic, respectively) were apparent, suggesting a relationship between these variables and diatom host–virus infection dynamics (Fig. 2d).

Given the prevalence of early and actively infected populations (C1 and C2) in Monterey Bay, we targeted this region to explore possible drivers of viral infection. The abundance of cell-associated diatom viruses was normalized to the total diatom transcript abundance to allow comparison of the degree of infection across samples with different diatom biomass loads (see Methods). PCA of the normalized cell-associated diatom virus abundance together with environmental metadata in Monterey Bay (Supplementary Table 1) yielded a biplot accounting for 74% of the variance, which depicted positive relationships between virus infection and biomass indicators (chlorophyll a, phaeophytin and $bSiO_2$) and negative relationships between virus infection and nutrient availability (N, P, dSi and Si stress; Supplementary Fig. 3). Pairwise linear regression analysis revealed that Si stress had the most robust correlation to the normalized abundance of cell-associated diatom viruses ($r^2 = 0.43$, $P = 0.02$; Supplementary Table 2), suggesting it was a significant predictor of

infection. No significant relationships with normalized cell-associated diatom virus abundance were apparent for the Point Reyes stations (Supplementary Table 2). However, taking both stations together, Si stress emerged as a driving factor for both the developing infection in Monterey Bay and the virus-induced mortality in Point Reyes stations (Fig. 2e). In Si-replete waters, low normalized abundance of cell-associated diatom viruses coupled with low ectoproteolytic enzyme activity and low phaeophytin suggested an early stage of infection. At moderate Si stress, the normalized cell-associated diatom virus abundance increased towards a maximum. Severely Si-stressed populations were characterized by decreased abundance of normalized cell-associated diatom virus, high ectoproteolytic enzyme activity and high phaeophytin, suggesting virus-mediated mortality. By integrating discrete diatom populations that span a wide range of Si stress, a relationship emerges that alludes to the temporal dynamics of viral infection, reconstructing early, active and lytic stages that are consistent with the regional characterization shown in Fig. 2a–d.

Given that diatoms take up Si and N at a 1:1 ratio[22] and low nitrate concentrations were concomitant with low dSi, we explored whether nitrate limitation also contributed to the observed diatom host–virus dynamics. We found no discernable relationship between the relative expression of diatom nitrate and ammonium transporters—proxies for nitrate limitation[23,24]—with cell-associated diatom virus abundance, further implicating Si limitation as a driver of diatom host–virus dynamics (Supplementary Fig. 4). To confirm a role for Si limitation in facilitating virus infection and host mortality, we conducted laboratory experiments with the centric, bloom-forming diatom, *Chaetoceros tenuissimus*, and two distinct host-specific viruses: CtenRNAV and CtenDNAV, ssRNA and ssDNA viruses, respectively. Infection progressed more rapidly in Si-limited cultures, which was indicated by an accelerated decline in host abundance and photosynthetic efficiency (F_v/F_m) and a larger increase in the relative proportion of dead cells compared with Si-replete infected cultures (Fig. 3a–d and Supplementary Fig. 5). Despite the clear impact on the timing of lysis, no significant difference was observed in the burst size (the number of viruses produced per host cell) between Si-replete and Si-limited cultures, indicating viral replication was not impacted by Si limitation (Fig. 3e). Notably, silicate was the primary limiting nutrient in Si-replete cultures at the time of host lysis (the remaining silicate in the media was $1.73 \pm 1.48\,\mu M$, whereas nitrate was an order of magnitude higher at $1.52 \pm 0.05\,mM$; $n = 9$). Together, these lab-based culture experiments demonstrate that Si limitation alone can serve as a driver of virus infection and mortality in diatoms.

Although the underlying mechanism remains unresolved, accelerated virus-induced mortality in Si-limited diatoms could be due to more efficient viral adsorption and entry due to reduced cellular $bSiO_2$. Alternatively, accelerated mortality may represent a strategy of the virus to ensure survival by hastening lysis when host physiology is compromised[25]. Si-limited cultures indeed exhibited elevated reactive oxygen species (Supplementary Fig. 6), an indicator of physiological stress and a signal in other phytoplankton thought to trigger host lysis[26].

Our findings demonstrate widespread signatures of virus infection in the California Current Ecosystem encompassing the early, active and lytic stages of infection. Together, these data illustrate the progression of infection along a gradient of Si stress throughout a diatom bloom (Fig. 4a). In Si-replete waters, diatoms grow rapidly and the early stages of viral infection are apparent. As the bloom progresses and Si becomes limiting, the population begins to experience moderate Si stress. Viral production becomes rampant and the infection begins to spread. Host physiological stress, associated with severe Si limitation, may serve as a signal to the virus (possibly through reactive oxygen species) that the bloom has reached peak host biomass and the loss of a viable host population is imminent.

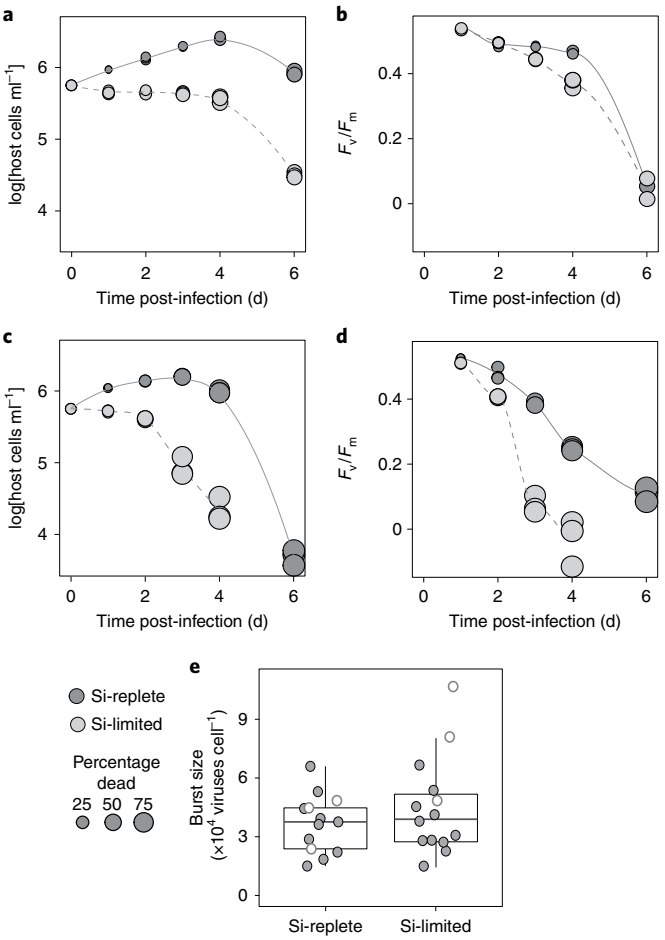


Fig. 3 | The impact of Si limitation on diatom host–virus dynamics in laboratory cultures of *C. tenuissimus*. a–d, Cultures were infected with either CtenRNAV (**a**,**b**) or CtenDNAV (**c**,**d**). Host abundance (**a**,**c**) and photosynthetic efficiency (**b**,**d**) in Si-replete (dark grey) and Si-limited cultures (light grey). The symbol size depicts the percentage of dead cells, as indicated by positive SYTOX staining. Each biologically independent sample ($n = 3$) is shown by individual symbols, with lines of best fit depicting a LOESS regression. Data are representative of one to four independent experiments. **e**, Box-and-whisker plot of the viral burst size (number of viruses produced per host) in biologically independent samples of infected Si-replete ($n = 13$) and Si-limited ($n = 14$) cultures for CtenRNAV (open symbols) and CtenDNAV (closed symbols). The boxes depict the median (horizontal line) and upper and lower quartiles of the data with whiskers encompassing data points within 1.5× of the interquartile range. Statistical significance between burst sizes was determined using Welch's two sample, two-sided *t*-test, which yielded *P* values of 0.13 and 0.56 for CtenRNAV and CtenDNAV, respectively.

This triggers mass virus-induced mortality and the bloom collapses as hosts lyse and release virions into the surrounding waters.

Our findings place virus infection of diatoms within the ecophysiological framework of Si availability and highlight the importance of the nutrient regime on host–virus dynamics, with critical implications for upper-ocean Si and C cycling (Fig. 4b). Accelerated infection and mortality in Si-limited populations would strengthen the 'viral shunt', coupling higher detrital loads with ectoproteolytic activity, and weaken the efficiency of the biological pump in diverse, Si-limited oceanic systems. As bacterial ectohydrolytic protease activity is a major driver of silica dissolution[27], virus-mediated mortality would enhance upper-ocean silica dissolution:production

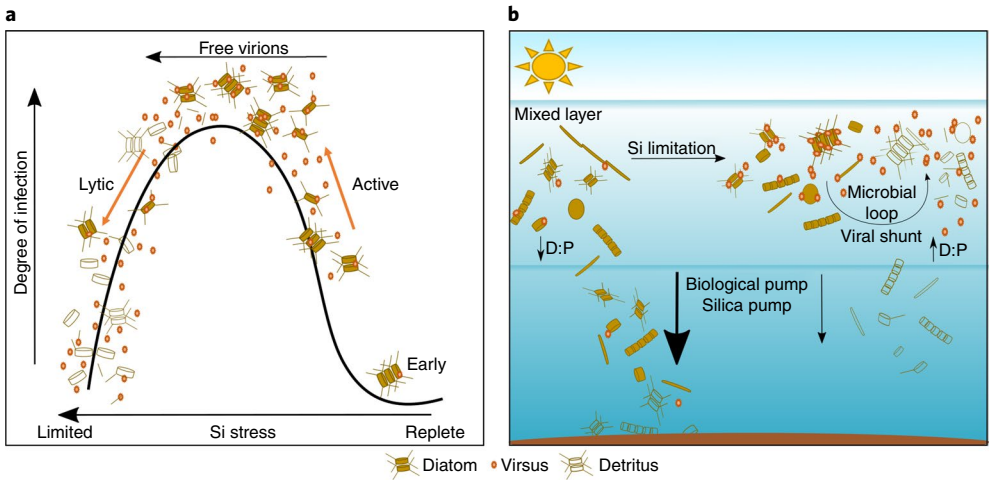


Fig. 4 | Conceptual model of diatom host–virus dynamics and impacts on biogeochemical cycling. a, Progression of virus infection across a gradient of Si stress during the course of a diatom bloom. The degree of infection increases from early infection in Si-replete waters to rampant viral infection as the developing bloom enters Si stress. Virus-induced mortality and bloom demise, associated with severe Si stress, occurs as the hosts lyse and release virions into the surrounding waters. **b**, Comparison of the biogeochemical consequences of diatom host–virus dynamics in replete and Si-limited populations. In replete populations, a low silica D:P ratio drives the biological and silica pumps, leading to the export of Si and C into the mesopelagic zone. Virus infection and mortality in Si-limited waters lead to an increase in D:P, weakening the biological and silica pumps while enhancing the supply of regenerative Si to resident diatoms (for example, uninfected cells, cells in the early stages of infection or resistant cells) through the stimulation of the microbial loop and viral shunt.

(D:P) ratios, thus weakening the silica pump[28,29] and enhancing the supply of regenerative Si to resident diatoms in the euphotic zone. High spatial and temporal variability in global silica D:P ratios has been largely attributed to changes in the composition of the bacterial community and the state of bloom demise[29,30]. We suggest that virus infection constitutes a heretofore unappreciated mechanistic control over diatom-mediated biogeochemical cycling in Si-limited regimes, promoting the retention of Si and C in the euphotic zone.

Methods

DYEatom cruise sampling and measurements. The DYEatom cruise was a multi-faceted study on diatom bloom dynamics and Si cycling conducted aboard the RV *Point Sur* (PS1312; 27 June 2013 to 5 July 2013; Chief Scientist: J. W. Krause). Given the dynamic currents and significant spatial variability, a surface-tethered drogue (Brightwaters Instrument Corporation) was deployed while on station and all hydrocasts were conducted within <100 m of the drogue to stay within the same water mass. Hydrocasts were conducted just after local sunrise. Water was collected using a rosette of 12 10-l Niskin bottles equipped with a Sea-Bird Electronics conductivity–temperature–depth sensor, Chelsea fluorometer and Bio-Spherical Instruments photosynthetically active radiation (PAR) sensor. The magnitude of PAR just below the surface (I_0) was used to determine the depths at which the PAR was 55%, 22%, 7% and 1% I_0; these depths matched the neutral density-screening chambers used to incubate samples in deck-board surface-water-cooled incubators.

The $bSiO_2$ production rates were measured using the radioisotope tracer ^{32}Si. Water was collected at each depth and 260 Bq $^{32}Si(OH)_4$ (>40 kBq μmol^{-1} Si) was added to 300 ml whole seawater and incubated at the respective light level in a deck-board incubator. Si uptake was measured after a 12 h incubation for all samples (except St7 samples, which were measured after 24 h due to the loss of the 12 h incubation bottles). The samples were filtered onto 1.2-μm pore-size polycarbonate filters (Millipore), dried on nylon planchettes and covered/secured in place with Mylar and a nylon ring. The activity of ^{32}Si was quantified using gas-flow proportional counting[31]. The gross $bSiO_2$ production rate (ρ) was normalized to $bSiO_2$ to determine the specific rate (V_b)[32]. Two methods were used to assess whether the ambient $Si(OH)_4$ concentration limited Si uptake; both used the same ^{32}Si additions, incubation conditions and processing described above. First, at St7, -8 and -11, eight 300-ml samples collected from 55% I_0 were manipulated to make an eight-point concentration gradient between ambient and +18.0 μM $Si(OH)_4$; the maximum concentration was assumed to saturate Si uptake. Silicon uptake has been shown to conform to a rectangular hyperbola described by the Michaelis–Menten equation:

$$V_b = \frac{V_{max} \times [Si(OH)_4]}{K_S + [Si(OH)_4]} \quad (1)$$

where V_{max} is the maximum specific uptake rate ($\rho \times bSiO_2^{-1}$) and K_S is the half-saturation constant, that is, the concentration where $V_b = 50\%$ of V_{max}. Data were fit to equation (1) using a nonlinear curve fitting algorithm (SigmaPlot 12.0). The second type of experiment used only two $[Si(OH)_4]$: ambient and +18.0 μM. The ratio of Si uptake at ambient $[Si(OH)_4]$ to +18.0 μM $[Si(OH)_4]$ defines the Si stress value (V_{amb}:V_{enh}). For 55% I_0 at St10, Si stress was assessed using gross Si-uptake rates (ρ, μmol l^{-1} d^{-1} Si) instead of specific uptake rates (normalized for $bSiO_2$) due to a compromised $bSiO_2$ sample at this depth. This two-point approach was conducted at all depths in the euphotic zone. Given the analytical error of the measurement, only Si stress values of 0–0.92 were considered analytically resolvable and used in our analysis. Values between 0.92 and 1.0 are considered within the error of detection[19].

Water for inorganic nutrients was syringe-filtered through a 0.6-μm polycarbonate filter (Millipore) into acid-washed plastic vials and immediately frozen at −20 °C. Nutrients (N, P) were quantified using an AutoAnalyzer[33] at UC Santa Barbara. Manual colorimetric analysis was used to measure dSi[34]. Seawater (200–350 ml) was filtered through a Millipore HAWP 0.45-μm pore-size cellulose filter for the determination of chlorophyll a using a 90% acetone extraction and acidification technique[33]. For $bSiO_2$, 600 ml seawater was filtered through a 1.2-μm pore-size polycarbonate filter (Millipore), immediately frozen at −20 °C and analysed using a modified sodium hydroxide digestion method[35]. Biomass for metatranscriptomic analysis was collected by filtration (after pre-filtration through a 200-μm nylon membrane) onto 47 mm, 1.2-μm pore-size polycarbonate filters at <5 psi for no longer than 15 min to minimize degradation. The filters were snap frozen in liquid N_2 and stored at −80 °C. Samples for free, extracellular virus analysis (~100 ml) were collected by syringe filtration onto 0.02-μm Anotop syringe filters (25 mm, Whatman) after pre-filtering through 0.22-μm Sterivex filters. The Anotops were sealed with parafilm and stored at −80 °C until processing.

Protease activity assays. The activity of ectoproteases in whole seawater was assessed using a fluorogenic substrate method modified for a microplate reader[36]. Aminopeptides labelled with the fluorogenic moiety, 7-amino-4-methylcoumarin (AMC), were used as substrates to assess the hydrolysis rates of serine, glycine and leucine. The enzymatic activity of each protease was analysed separately. Triplicate 200 μl reactions of each fluorogenic substrate and seawater were prepared in a 96-well plate. The microplates were incubated in the dark at in situ temperature for 24 h. The fluorescence of each reaction was measured approximately every 4 h throughout the incubation. A standard curve of AMC fluorescence was used to convert fluorescence to the amount of substrate (μmol) that was enzymatically cleaved per litre of seawater.

RNA extraction, sequencing and assembly. One billion copies of each of two commercially available RNA standards (purified RNA transcripts) were added to each filter (nos 1 and 8, ArrayControl Spots and Spikes, Thermo Fisher Scientific; with polyA tails). RNA was extracted using TRIzol reagent (Life Technologies).

For each sample, complementary DNA (300–18,500 pg µl^{-1} concentration range) was prepared from 50 ng total RNA using the SuperScript III first strand synthesis system with random hexamers (Life Technologies) and then amplified using an AccuPrime PCR system kit (Life Technologies) in a reaction containing 1 µl cDNA, 1×AccuPrime buffer II, 0.75 U AccuPrime Taq high fidelity and a 200 nM final primer concentration alongside a no-template control. Amplification was performed as follows: an initial denaturation at 95 °C for 2 min, followed by 30 cycles of 95 °C for 20 s, 56 °C for 30 s and 72 °C for 5 min. The products (2 µl sample, 5 µl no-template control) were run on a 1% agarose gel, cleaned with Ampure XP beads (Beckman Coulter) and resuspended in 25 µl Qiagen elution buffer. The samples were visualized on an agarose gel and quantified using the PicoGreen Quant-IT assay (Life Technologies); 45 ng of 18 S amplicons were pooled for 454 pyrosequencing.

Ribosomal RNA (18S) was sequenced using 454 pyrosequencing. TAReuk454FWD1 (5′-CCAGCASCYGCGGTAATTCC-3′) and TAReukREV3 (5′-ACTTTCGTTCTTGATYRA-3′) were used to target a 500 bp region of the v4 region[37]. FLX Titanium adaptors (A-adaptor sequence: 5′-127-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3′; B-adaptor sequence: 5′-128 CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3′) and 10 bp multiplex identifier barcodes were used for multiplexed sequencing.

Library quality control, emulsion PCR, enrichment and 454 sequencing was done according to the manufacturer's protocol (Roche Diagnostics) with the following modifications: the KAPA Biosystems library quantification kit for qPCR was used to estimate the number of molecules needed for emulsion PCR, automation (BioMek FX; Beckman Coulter) was used to 'break' the emulsions after emulsion PCR and butanol was used for ease of handling during the breaking process. Bead enrichment was automated with a Robotic Enrichment Module (Roche Diagnostics).

The 18 S rRNA 454 reads were demultiplexed using the sff file utility from Roche/454 and converted from a standard flowgram to Fasta format using sff2fastq, followed by primer removal, quality control, trimming and dereplication. Chimeric sequences were removed using USEARCH[38], reads were trimmed to a quality score of ten over a two-base window, operational taxonomic units were clustered using SWARM[39] and classified using FASTA36 from the FASTA package. Taxonomic annotations were assigned using GLSEARCH36 (ref. [40]) with a modified PR2 database with updates from Tara Oceans W2 (ref. [41]).

Metatranscriptome libraries were constructed using 500 ng total RNA and a TruSeq RNA sample preparation kit (Illumina) following the low-throughput protocol. The mean size of the final libraries was confirmed to be between 359 and 420 base pairs (bp) using an Agilent Bioanalyzer 2100. The libraries were paired-end sequenced (2×150 bp) on the Illumina HiSeq platform.

Reads were trimmed to remove primers and areas of low sequence quality (that is, <30 bp long and a quality score <33). Illumina reads were paired. The rRNA reads were removed using Ribopicker version 0.4.3 (ref. [42]). Large fraction reads were assembled using CLC Genomics Workbench 9.5.3 first by library, then overall. Ab initio open-reading-frame (ORF) prediction was performed using FragGeneScan[43]. The ORFs were screened for contamination in the form of rRNA, internal transcribed spacers and primers. Open reading frames with closer homology to a known organelle gene than to a nuclear gene in the same reference organism were filtered.

The ORFs were annotated via BLASTP[44] alignment (E-value < 10^{-3}) to a comprehensive protein database, phyloDB, as well as screened for function de novo by assigning Pfams, TIGRfams and transmembrane tmHMMs with hmmer 3.0 (http://hmmer.org/)[45]. PhyloDB 1.076 consists of 24,509,327 peptides from 19,962 viral, 230 archaeal, 4,910 bacterial and 894 eukaryotic taxa. It includes peptides from KEGG, GenBank, JGI, ENSEMBL, CAMERA and various other repositories, as well as from the 410 taxa of the Marine Microbial Eukaryotic Transcriptome Sequencing Project[46]. Taxonomic annotation of ORFs was also conducted via BLASTP to phyloDB.

Added RNA spikes were used to estimate the number of copies per litre within each library[47–50]. Reads per bp for the ORFs of interest were divided by the spike reads recovered per bp per copy spike added per litre of seawater for each library. Recovery of spike 8 (2,000 bp without tail) was used for ORFs > 1,000 bp. For ORFs < 1,000 bp, spike 1 (750 bp without tail) was used. Diatom virus contigs were identified using a homology-based query (BLASTP, E-value < 10^{-3}) with RdRP and replicase genes from known ssRNA (genus *Bacillarnavirus*) and ssDNA (genus *Bacilladnavirus*) diatom viruses, respectively. No ssDNA viruses were identified. Detailed annotation, assembled contigs and virus abundance data are available in Supplementary Data 1. Cluster analysis, PCA, linear regression analysis and other statistical analyses were performed in R version 3.5.1 'Feather Spray'. Before PCA, data were scaled using a correlation matrix to standardize each of the variables used in the analysis. The optimal number of clusters in K-means analysis was determined using the elbow method.

To assess the degree of infection across samples with different diatom biomass loads, the total diatom virus RdRP reads from the metatranscriptomes were normalized to the total diatom reads, analogous to transcript per million normalization[51] used for comparing relative gene expression between samples and similar to group normalization methods used in previous studies[24,52–54].

Phylogenetic analysis. Reference RdRP amino acid sequences were obtained from the National Center for Biotechnology Information (NCBI), trimmed and aligned using ClustalW[55] in Mega 7 (ref. [56]). A maximum-likelihood tree was constructed from the reference alignment using RAxML version 8.2.10 with the PROTGAMMAWAGF substitution model and 100 bootstrap replicates[57]. Metatranscriptomic reads from the same contig were concatenated and aligned to the reference alignment using hmmalign and the HMM profile of the reference alignment (built using hmmbuild[46]). Tree files were packaged using taxtastic version 0.8.3 and environmental contigs were placed on the reference tree using pplacer version 1.1.alpha19 (ref. [58]). The tree was exported and the posterior probability was calculated using guppy version 1.1.

Anotop nucleic acid extraction and quantitative reverse-transcription PCR. Anotop syringe filters were thawed at room temperature and subjected to nucleic acid extraction using a MasterPure complete DNA and RNA extraction kit (Epicentre), according to Mueller et al.[59] with minor modifications. Briefly, 500 µl of 1×Tissue and cell lysis buffer with 100 µg ml^{-1} proteinase K was injected into the Anotop, sealed at both ends with Luer-Lok syringes and held in place with tape to ensure the buffer remained in contact with the filter. The entire assembly was then incubated in a rotisserie oven for 15 min at 65 °C, rotating at 10 r.p.m. The lysate buffer was recovered by drawing into one of the syringes and transferring to a microfuge tube, chilled on ice for 5 min and treated with MPC protein precipitation reagent according to the MasterPure protocol. An additional centrifugation step after protein precipitation was done to ensure the complete removal of precipitate. Nucleic acids were then precipitated with isopropanol at −20 °C for 1 h. After a 20 min centrifugation at maximum speed at 4 °C, the pellet was washed with cold 70% ethanol and resusupended in 35 µl TE buffer (pH 8). The samples were treated with TURBO DNAse (Invitrogen) according to the manufacturer's protocol. The RNA was quantified using a Nanodrop 1000 (Thermo Scientific) and 50–180 ng was reverse transcribed using the SuperScript III First Strand Synthesis system with random hexamers (Invitrogen).

Free, extracellular diatom viruses were targeted by designing primers against each of the three identified diatom virus contigs in the metatranscriptomes. To increase the signal and specificity, a set of outer and nested primer pairs were designed for use in a two-round PCR reaction (Supplementary Table 3). This approach is routinely used in virology to increase both the sensitivity and specificity when quantifying viruses that are low in abundance[60–62]. In Round 1, 17 ng cDNA (based on the RNA concentration used in the reverse transcription reaction and assuming 100% efficiency) was used as the template with the outer set of primers (500 nM of each primer), 1.25 U Taq DNA Polymerase (Invitrogen), 1×PCR buffer, 0.2 mM dNTPs and 2 mM $MgCl_2$ in the following reaction performed in an Eppendorf Mastercycler EP S thermal cycler: initial denaturation at 95 °C for 2 min, 20 cycles of 95 °C for 30 s, 53–54 °C (Supplementary Table 3) for 30 s and 72 °C for 45 s, followed by 72 °C for 7 min. For the qRT–PCR, 1 µl of a 1:20 dilution (in TE buffer, pH 8) of the Round 1 product was used as the template in a 10 µl reaction using PowerUp SYBR green master mix (Applied Biosystems) and the nested or semi-nested set of primers (500 nM each). The following reaction was performed in a Mx3000P qPCR system (Strategene): 50 °C for 2 min 95 °C for 2 min, and 40 cycles of 95 °C for 30 s, 52–54 °C (Supplementary Table 3) for 30 s and 72 °C for 45 s followed by melt curve analysis to check for primer dimers. The amplified products (112–204 bp and 72–129 bp for the outer and nested products, respectively) were subsequently run on a 1.5% agarose gel to verify both the size and specificity of the PCR reactions. Standard curves for gene-copy quantification and primer efficiencies were generated using gel-purified (PureLink Quick Gel) amplicons (after two rounds of PCR with the outer primers) and serial dilutions in qRT–PCR reactions to obtain 1–1,000 copies of purified product. Extracellular-virus abundance (copies l^{-1}) was calculated by dividing the number of copies in the qRT–PCR reaction by the amount of RNA used in the reaction (17 ng), then multiplying this by the total amount of RNA extracted from the Anotop to obtain the number of copies on the Anotop. This was then divided by the total volume of seawater filtered.

Culturing conditions and infection experiments. *C. tenuissimus* Meunier and its associated viruses isolated from the coastal waters of Japan[63] were provided by Y. Tomaru. Host cultures were maintained in SWM-3 media[64] with 0.2 mM silicate, 2 mM nitrate and 2 nM Na_2SeO_3(ref. [63]) at 15 °C on a 12:12 light:dark cycle at ~120 µmol photons m^{-2} s^{-1}. All culture work was performed in polycarbonate culture flasks to minimize Si contamination. Cell abundance was measured using flow cytometry (BD Accuri C6; 488 nm excitation, >670 nm emission) at a flow rate of 14 µl min^{-1} with a 10-µm core diameter. Cell populations were identified by plotting the chlorophyll fluorescence and forward scatter (Supplementary Fig. 7).

For the Si-limitation experiments, exponential phase cultures growing in Si-replete media were centrifuged at 4,000g for 10 min, washed in SWM-3 media without added Si (SWM − Si) and resuspended in either replete or SWM − Si media. The cultures were grown for 3–4 d until Si limitation was apparent from the differences in specific growth rates between the Si-replete and Si-limited cultures. The cultures were then subdivided and one set was infected with either CtenDNAV[65] or CtenRNAV[53] at a multiplicity of infection of 10. The F_v/F_m values were measured using a custom-built Fluorescence Induction and Relaxation System[66] providing measurements of F_o (minimum) and F_m (maximum) fluorescence yields. The maximum efficiency of the photosystem

II was calculated as $F_v/F_m = (F_m - F_o)/F_m$. The cultures were low-light acclimated for 15 min before measurements. The relative proportion of live/dead cells in cultures was determined by staining with SYTOX green (Thermo Fischer) at a final concentration of 1 µM for 15 min in the dark at room temperature and subsequent analysis using flow cytometry (BD Accuri C6; 488 nm excitation, 533/30 emission). Unstained samples were used to set a threshold for background green fluorescence. The relative number of dead cells in a population (percent positively stained) was determined as those that exceeded the background fluorescence of the unstained sample (Supplementary Fig. 7). Intracellular reactive oxygen species were quantified by staining with 2′,7′-dichlorodihydrofluorescein diacetate (CM-H_2DCFDA; Thermo Fisher Scientific) at a final concentration of 5 µM for 1 h in the dark at room temperature and subsequent analysis using flow cytometry (BD Accuri C6; 488 nm excitation, 533/30 emission). The mean background cellular fluorescence of unstained samples (relative fluorescence units per cell) was subtracted from the mean fluorescence of positively stained populations. Nitrate, dSi and $bSiO_2$ were measured using standard protocols[33–35].

Virus abundance was measured in lysates generated by filtering cultures through a 0.22-µm pore-size filter to remove the cellular debris. To confirm there was minimal loss of viruses after the 0.22-µm pre-filtration step, we measured the virus abundance after sequential filtration through a 1.2, 0.22 and 0.02-µm pore-size filter. There was no difference, within the 95% confidence levels of the measurement, between the 1.2 and 0.22-µm lysates (9.4×10^9 and 1.4×10^{10} viruses ml^{-1}, respectively). Notably, there was a 98% reduction in virus abundance following filtration through a 0.02-µm pore-size filter, demonstrating the efficient retention of diatom viruses on filters with this pore size (as expected given the 20–40 nm diameter of diatom viruses). The lysates were stored at 4 °C and processed within one month of collection. The abundance of viruses was measured using the most-probable-number assay[67]. Unlike methods that quantify the total number of virus particles (infectious and non-infectious viruses), such as flow cytometry and fluorescent stains[68], the most-probable-number assay only quantifies infectious virus particles. Briefly, in a 96-well plate, exponential-phase cultures at 5×10^5 cells ml^{-1} were incubated with ten-fold serial dilutions of each viral lysate. Eight replicate infections for each lysate dilution were performed. The plates were sealed with parafilm and incubated at standard growth conditions with top lighting and monitored for cell lysis. Final scores were obtained after 13 d when lysis was complete. Virus abundance (infectious units per ml) was calculated using the EPA-MPN calculator[69]. Burst size was calculated by dividing the final virus abundance by the decrease in host abundance.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All cruise-related data are available publicly at the Biological and Chemical Oceanography Data Management Office under the project number 550825 (https://www.bco-dmo.org/project/550825). The metatranscriptomic data reported in this paper have been deposited in the NCBI sequence read archive (BioProject accession no. PRJNA528986, BioSample accession nos SAMN11263616–SAMN11263639 and SAMN11258802–SAMN11258825). The assembled contigs used in this study can also be found at https://scripps.ucsd.edu/labs/aallen/data/ and BCO-DMO project number 558198 (https://www.bco-dmo.org/project/558198). All data generated or analysed during the current study are included in this published article and its supplementary information files.

Received: 11 January 2019; Accepted: 30 May 2019;
Published online: 15 July 2019

References

1. Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A. & Quéguiner, B. Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochem. Cycles* **9**, 359–372 (1995).
2. Malviya, S. et al. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl Acad. Sci. USA* **113**, E1516–E1525 (2016).
3. Jin, X., Gruber, N., Dunne, J. P., Sarmiento, J. L. & Armstrong, R. A. Diagnosing the contribution of phytoplankton functional groups to the production and export of particulate organic carbon, $CaCO_3$, and opal from global nutrient and alkalinity distributions. *Global Biogeochem. Cycles* **20**, GB2015 (2006).
4. Smetacek, V. et al. Deep carbon export from a Southern Ocean iron-fertilized diatom bloom. *Nature* **487**, 313–319 (2012).
5. Agusti, S. et al. Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. *Nat. Commun.* **6**, 7608 (2015).
6. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999).
7. Suttle, C. A. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
8. Dugdale, R. C. & Wilkerson, F. P. Silicate regulation of new production in the equatorial Pacific upwelling. *Nature* **391**, 270–273 (1998).
9. Krause, J. W. et al. The effects of biogenic silica detritus, zooplankton grazing, and diatom size structure on silicon cycling in the euphotic zone of the eastern equatorial Pacific. *Limnol. Oceanogr.* **55**, 2608–2622 (2010).
10. Tréguer, P. J. & De La Rocha, C. L. The world ocean silica cycle. *Ann. Rev. Mar. Sci.* **5**, 477–501 (2013).
11. Wilhelm, S. W. & Suttle, C. A. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* **49**, 781–788 (1999).
12. Laber, C. P. et al. *Coccolithovirus* facilitation of carbon export in the North Atlantic. *Nat. Microbiol.* **3**, 537–547 (2018).
13. Brussaard, C. P. D. Viral control of phytoplankton populations—a review. *J. Eukaryot. Microbiol.* **51**, 125–138 (2004).
14. Nagasaki, K. Dinoflagellates, diatoms, and their viruses. *J. Microbiol.* **46**, 235–243 (2008).
15. Steward, G. F. et al. Are we missing half of the viruses in the ocean? *ISME J.* **7**, 672–679 (2013).
16. Culley, A. New insight into the RNA aquatic virosphere via viromics. *Virus Res.* **244**, 84–89 (2018).
17. Moniruzzaman, M. et al. Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. *Nat. Commun.* **8**, 16054 (2017).
18. Lassiter, A. M., Wilkerson, F. P., Dugdale, R. C. & Hogue, V. E. Phytoplankton assemblages in the CoOP-WEST coastal upwelling area. *Deep Sea Res. II* **53**, 3063–3077 (2006).
19. Krause, J. W., Brzezinski, M. A., Villareal, T. A. & Wilson, C. Increased kinetic efficiency for silicic acid uptake as a driver of summer diatom blooms in the North Pacific subtropical gyre. *Limnol. Oceanogr.* **57**, 1084–1098 (2012).
20. Smith, D. C., Steward, G. F., Long, R. A. & Azam, F. Bacterial mediation of carbon fluxes during a diatom bloom in a mesocosm. *Deep Sea Res. II* **42**, 75–97 (1995).
21. Berges, J. A. & Falkowski, P. G. Physiological stress and cell death in marine phytoplankton: induction of proteases in response to nitrogen or light limitation. *Limnol. Oceanogr.* **43**, 129–135 (1998).
22. Brzezinski, M. A. The Si:C:N ratio of marine diatoms: interspecific variability and the effect of some environmental variables. *J. Phycol.* **21**, 347–357 (1985).
23. Bender, S. J., Durkin, C. A., Berthiaume, C. T., Morales, R. L. M. & Armbrust, E. Transcriptional responses of three model diatoms to nitrate limitation of growth. *Front. Mar. Sci.* **1**, 3 (2014).
24. Lampe, R. H. et al. Divergent gene expression among phytoplankton taxa in response to upwelling. *Environ. Microbiol.* **20**, 3069–3082 (2018).
25. Wang, N. Lysis timing and bacteriophage fitness. *Genetics* **172**, 17–26 (2006).
26. Sheyn, U., Rosenwasser, S., Ben-Dor, S., Porat, Z. & Vardi, A. Modulation of host ROS metabolism is essential for viral infection of a bloom-forming coccolithophore in the ocean. *ISME J.* **10**, 1742 (2016).
27. Bidle, K. D. & Azam, F. Accelerated dissolution of diatom silica by marine bacterial assemblages. *Nature* **397**, 508–512 (1999).
28. Dugdale, R. C., Wilkerson, F. P. & Minas, H. J. The role of a silicate pump in driving new production. *Deep Sea Res. I* **42**, 697–719 (1995).
29. Bidle, K. D., Brzezinski, M. A., Long, R. A., Jones, J. L. & Azam, F. Diminished efficiency in the oceanic silica pump caused by bacterially-mediated silica dissolution. *Limnol. Oceanogr.* **48**, 1855–1868 (2003).
30. Brzezinski, M. A., Jones, J. L., Bidle, K. D. & Azam, F. The balance between silica production and silica dissolution in the sea: insights from Monterey Bay, California applied to the global data set. *Limnol. Oceanogr.* **48**, 1846–1854 (2003).
31. Krause, J. W., Nelson, D. M. & Brzezinski, M. A. Biogenic silica production and the diatom contribution to primary production and nitrate uptake in the eastern equatorial Pacific Ocean. *Deep Sea Res. II* **58**, 434–448 (2011).
32. Brzezinski, M. A. & Phillips, D. R. Evaluation of ^{32}Si as a tracer for measuring silica production rates in marine waters. *Limnol. Oceanogr.* **42**, 856–865 (1997).
33. Brzezinski, M. A. & Washburn, L. Phytoplankton primary productivity in the Santa Barbara Channel: effects of wind-driven upwelling and mesoscale eddies. *J. Geophys. Res. Ocean.* **116**, C12013 (2011).
34. Nelson, D. M. & Brzezinski, M. A. Diatom growth and productivity in an oligo-trophic midocean gyre: a 3-yr record from the Sargasso Sea near Bermuda. *Limnol. Oceanogr.* **42**, 473–486 (1997).
35. Krause, J. W., Lomas, M. W. & Nelson, D. M. Biogenic silica at the Bermuda Atlantic Time-series Study site in the Sargasso Sea: Temporal changes and their inferred controls based on a 15-year record. *Glob. Biogeochem. Cycles* **23**, 1–14 (2009).
36. Hoppe, H. in *Handbook of Methods in Aquatic Microbial Ecology* 1st edn (eds Kemp, P. F. et al.) 423–431 (Lewis Publishers, 1993).
37. Stoeck, T. et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).
38. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

39. Mahé, F., Rognes, T., Quince, C., de Vargas, C. & Dunthorn, M. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* **3**, e1420 (2015).
40. Pearson, W. R. Finding protein and nucleotide similarities with FASTA. *Curr. Protoc. Bioinform.* **53**, 3.9.1–25 (2016).
41. De Vargas, C. et al. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).
42. Schmieder, R., Lim, Y. W. & Edwards, R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28**, 433–435 (2011).
43. Rho, M., Tang, H. & Ye, Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**, e191 (2010).
44. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
45. Sonnhammer, E. L. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. In *Proc. 6th International Conference on Intelligent Systems for Molecular Biology* (eds Glasgow, J. et al.) 175– 182 (Association for the Advancement of Artificial Intelligence, 1998).
46. Keeling, P. J. et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).
47. Gifford, S. M., Sharma, S., Rinta-Kanto, J. M. & Moran, M. A. Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME J.* **5**, 461–472 (2011).
48. Moran, M. A. et al. Sizing up metatranscriptomics. *ISME J.* **7**, 237–243 (2013).
49. Satinsky, B. M., Gifford, S. M., Crump, B. C. & Moran, M. A. in *Methods in Enzymology* Vol. 531 (Ed Delong, E. F.) 237–250 (Elsevier, 2013).
50. Bertrand, E. M. et al. Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl Acad. Sci USA* **112**, 9938–9943 (2015).
51. Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**, 281–285 (2012).
52. Ottesen, E. A. et al. Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *ISME J.* **5**, 1881–1895 (2011).
53. Marchetti, A. et al. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl Acad. Sci. USA* **109**, E317–E325 (2012).
54. Alexander, H., Jenkins, B. D., Rynearson, T. A. & Dyhrman, S. T. Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proc. Natl Acad. Sci. USA* **112**, E2182–E2190 (2015).
55. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
56. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
57. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
58. Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinform.* **11**, 538 (2010).
59. Mueller, J. A., Culley, A. I. & Steward, G. F. Variables influencing extraction of nucleic acids from microbial plankton (viruses, bacteria, and protists) collected on nanoporous aluminum oxide filters. *Appl. Environ. Microbiol.* **80**, 3930–3942 (2014).
60. Carriere, M. et al. A novel, sensitive, and specific RT-PCR technique for quantitation of hepatitis C virus replication. *J. Med. Virol.* **79**, 155–160 (2007).
61. Pasternak, A. O. et al. Highly sensitive methods based on seminested real-time reverse transcription-PCR for quantitation of human immunodeficiency virus type 1 unspliced and multiply spliced RNA and proviral DNA. *J. Clin. Microbiol.* **46**, 2206–2211 (2008).
62. Hernández-Arteaga, S. & López-Revilla, R. Ultrasensitive quantitation of human papillomavirus type 16 E6 oncogene sequences by nested real time PCR. *Infect. Agents Cancer* **5**, 9 (2010).
63. Shirai, Y. et al. Isolation and characterization of a single-stranded RNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus* Meunier. *Appl. Environ. Microbiol.* **74**, 4022–4027 (2008).
64. Chen, L., Edelstein, T. & McLachlan, J. *Bonnemaisonia hamifera* hariot in nature and in culture. *J. Phycol.* **5**, 211–220 (1969).
65. Tomaru, Y., Shirai, Y., Toyoda, K. & Nagasaki, K. Isolation and characterisation of a single-stranded DNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus*. *Aquat. Microb. Ecol.* **64**, 175–184 (2011).
66. Gorbunov, M. Y. & Falkowski, P. G. Fluorescence induction and relaxation (FIRe) technique and instrumentation for monitoring photosynthetic processes and primary production in aquatic ecosystems. In *Photosynthesis: Fundamental Aspects to Global Perspectives-Proc. 13th International Congress of Photosynthesis* (eds Van der Est, A. & Bruce, D.) 1029–1031 (Allen and Unwin, London, 2004).
67. Suttle, C. A. in *Handbook of Methods in Aquatic Microbial Ecology* 1st edn (eds Kemp, P. F. et al.) 121–134 (Lewis Publishers, 1993).
68. Lippé, R. Flow virometry: a powerful tool to functionally characterize viruses. *J. Virol.* **92**, e01765–17 (2018).
69. Klee, A. J. A computer program for the determination of most probable number and its confidence limits. *J. Microbiol. Methods* **18**, 91–98 (1993).

Acknowledgements

We would like to thank the captain and the crew of the RV *Point Sur*, J. Jones, H. McNair, E. Lachenmyer, I. Marquez and J. Ossolinski, for technical assistance during the cruise. Surface-tethered drogues used on the cruise were provided by R. Chant and E. Hunter. Thank you to Y. Tomaru for providing the laboratory diatom host–virus systems; J. Latham for technical support; B. Knowles, E. Zelzion and K. Bondoc for their useful discussions on statistical analysis; and B. Knowles and J. Nissimov for their helpful comments on the manuscript. This work was supported by grants from the National Science Foundation (grant nos OCE-13339329 and OCE-1559179 to K.T., OCE-1334387 to M.A.B., OCE-1155663 to J.W.K., and OCE-1637632 and OCE-1756884 to A.E.A.), the Gordon and Betty Moore Foundation (grant nos GBMF3301 to B.A.S.V.M. and K.D.B., GBMF3789 to K.D.B. and GBMF3828 to A.E.A.), the National Oceanic and Atmospheric Administration (grant no. NA15OAR4320071 to A.E.A.) and a postdoctoral fellowship from the Simons Foundation (grant no. 548156 to C.F.K.). Salary support for C.F.K. was also provided by the Institute of Earth, Ocean and Atmospheric Sciences at Rutgers University, the Rappaport Fund for Advanced Studies and Israel's Council for Higher Education.

Author contributions

C.F.K. and K.T. conceived the project, designed the experiments and wrote the paper. C.F.K. and W.P.B. conducted the laboratory culture-based experiments. C.F.K. performed the metatranscriptomic and statistical analyses. C.F.K. and K.T. processed and analysed the field samples for extracellular virus. M.M. assisted with the 18S rRNA and RdRP phylogenetic analyses. J.W.K. was the Chief Scientist of the DYEatom cruise. J.W.K., M.A.B., B.A.S.V.M., K.D.B. and K.T. were involved in the cruise planning. J.W.K. and M.A.B. collected and analysed the silica-production, nutrient and bulk particle data. B.R.E. and B.A.S.V.M. conducted and provided the on-ship protease activity data. K.D.B. and K.T. collected all other field samples. A.E.A. extracted the RNA and generated the metatranscriptome and 18S rRNA data. A.E.A. and J.P.M. performed the bioinformatic analyses. All authors provided comments on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0502-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.T.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Kimberlee Thamatrakoln

Last updated by author(s): May 24, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	BD Accuri C6 Software; SoftMax Pro 6.2.1
Data analysis	R 3.5.1 "Feather Spray", Microsoft Excel 2010, BLAST+ 2.2.31, HMMer 3.0, SigmaPlot 12.0, Mega 7, RAxML v8.2.10, HMMER 3.2.1 (hmmalign and hmmbuild), taxtastic v.0.8.3, pplacer v.1.1.alpha19, guppy v.1.1, USEARCH v8.0.1616, SWARM v2.1.10, GLSEARCH36, Ribopicker v.0.4.3, FragGeneScan v1.16, EPA-MPN Calculator 2.0, sff2fastq v0.9.2, FASTA36 v36.3.6, Blastall v2.2.26, CLC Genomics Workbench 9.5.3

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All cruise related data are available publicly at the Biological & Chemical Oceanography Data Management Office under project numbers 550825 (https://www.bco-dmo.org/project/550825). The metatranscriptomic data reported in this paper have been deposited in the NCBI sequence read archive (BioProject accession no. PRJNA528986: BioSample accession nos. SAMN11263616 - SAMN11263639 and SAMN11258802-SAMN11258825). Assembled contigs used in this study can also be found at https://scripps.ucsd.edu/labs/aallen/data/ and BCO-DMO project number 558198 (https://www.bco-dmo.org/project/558198). All data generated or analyzed during the current study are included in this published article and its supplementary information files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This was a multi-faceted study on diatom bloom dynamics, virus infection, and silicon cycling conducted aboard the R/V Point Sur in the California Coastal Ecosystem and in laboratory cultures of a model diatom host-virus system. There were no treatments, design structure or experimental units associated with observational field samples. Two treatment factors were used for culture experiments: infected or uninfected, replete or silicon-limiting media. A nested design structure with repeated measures for each experimental unit was used to test the interaction between nutrient concentration and viral infection. Within an experiment, there were 12 experimental units, with triplicates for each treatment.
Research sample	Samples from natural populations consisted of whole seawater collected from multiple depths and contained a community of organisms. Laboratory experiments were conducted using the diatom, Chaetoceros tenuissimus and its associated viruses, CtenDNAV and CtenRNAV, provided by Dr. Yuji Tomaru at the National Research Institute of Fisheries and Environment of Island Sea in Japan.
Sampling strategy	Field samples were collected based on the surface irradiance within the euphotic zone. The duration of the cruise and the time available for each station dictated the sample size and was deemed sufficient based on previously published studies. For laboratory culture experiments, sample size was determined to be sufficient based on low variability between replicates.
Data collection	Field based samples were collected using Niskin bottles attached to a rossette equipped with a conductivity, temperature and depth sensor. Data were recorded at the time of collection on log sheets. Laboratory culture data were collected either via centrifugation or filtration and recorded in notebooks.
Timing and spatial scale	Field-based sampling was conducted 27 June- 5 July 2013 sampled over a ~32,000 sq km region near Monterey Bay and Point Reyes, CA. Stations were sampled daily because that was the time needed to sample and process a station and proceed to another station. Periodicity of sampling in laboratory cultures is indicated on the x-axis of the associated figure. Samples were generally collected daily with gaps due to non-working days. This periodicity was sufficient to resolved changes and differences throughout the approximately 1 week duration of the experiment.
Data exclusions	No data were excluded.
Reproducibility	Field based measurements were not reproduced as these represent observational snapshots in time and space. For culture-based laboratory experiments, reproducibility was confirmed across independent experiments with biologically independent diatom cultures and virus stocks generated separately for each experiment. Detailed methods of experimental design and sampling are reported to ensure reproducibility.
Randomization	Randomization is not relevant to the field-based aspect of this study because they represent observational samples. Station locations were chosen based on in-situ chlorophyll fluorescence but the exact water mass that was sampled was arbitrary. For laboratory culture work, cultures were subdivided into replicate flasks and those to which viruses were added were chosen randomly.
Blinding	Except for the use of in-situ chlorophyll fluorescence measurements, we had no prior knowledge of bloom state prior to sampling. Samples were collected with an alpha-numeric sequential designation and were analyzed without knowledge of the station or depth the sample was derived from.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Sampling took place during the early summer and conditions were typical for that region.
Location	Oceanographic sampling took place between 36-39 degrees North and 121-123 degrees West. Samples were collected throughout the euphotic zone ranging from the surface down to 55 meters.
Access and import/export	Sampling locations were accessed while on-board the R/V Point Sur and compliance with all relevant laws was ensured by the Captain and the crew operating the vessel.
Disturbance	No disturbance was caused by this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Chaetoceros tenuissiumus was provided by Dr. Yuji Tomaru at the National Research Institute of Fisheries and Environment of Island Sea in Japan.
Authentication	18S rDNA sequencing
Mycoplasma contamination	Cell lines were not tested for Mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Throughout each experiment, cultures were sampled in a biological hood and measured immediately for abundance in a 96-well plate in the BD Accuri C6 flow cytometer as described in the methods. For staining procedures, samples were stained and measured immediately following the incubation with the stain. In order to insure similar incubations times between staining and sample measurement for all samples, staining was staggered accordingly.
Instrument	BD Accuri C6
Software	Data was collected and analyzed using the BD Accuri C6 software.
Cell population abundance	Cell abundance was measured using flow cytometry (BD Accuri C6; 488 nm excitation, >670 nm emission) at a flow rate of 14 µL min-1 and a 10 µm core diameter. Cell populations were identified by plotting chlorophyll fluorescence and forward scatter.
Gating strategy	For live/dead staining, unstained samples were used to set a threshold for background green fluorescence (488 nm excitation, 533/30 emission). The relative number of dead cells in a population (percent positively stained, with preliminary gating of the population determined by forward scatter) was determined as those that exceeded the background fluorescence of the unstained sample. Intracellular reactive oxygen species (488 nm excitation, 533/30 emission) was measured by subtracting the mean fluorescence of positively stained populations from the mean background cellular fluorescence of unstained samples (RFU event-1). Threshold of unstained population was set with a preliminary gate of the population identified with chlorophyll fluorescence (488 nm excitation, >670 nm emission) and forward scatter.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

nature microbiology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41564-019-0502-x

In the format provided by the authors and unedited.

Silicon limitation facilitates virus infection and mortality of marine diatoms

Chana F. Kranzler[1], Jeffrey W. Krause[2,3], Mark A. Brzezinski[4], Bethanie R. Edwards[5], William P. Biggs[1], Michael Maniscalco[4], John P. McCrow[6], Benjamin A. S. Van Mooy[7], Kay D. Bidle[1], Andrew E. Allen[6,8] and Kimberlee Thamatrakoln[1]*

[1]Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA. [2]Dauphin Island Sea Lab, Dauphin Island, AL, USA. [3]Department of Marine Sciences, University of South Alabama, Mobile, AL, USA. [4]Marine Science Institute, University of California, Santa Barbara, CA, USA. [5]Department of Earth and Planetary Science, University of California, Berkeley, CA, USA. [6]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA, USA. [7]Woods Hole Oceanographic Institution, Woods Hole, MA, USA. [8]Scripps Institution of Oceanography, University of California, San Diego, CA, USA. *e-mail: thamat@marine.rutgers.edu

Supplementary Information

Silicon limitation facilitates virus infection and mortality of marine diatoms

Chana F. Kranzler[1], Jeffrey W. Krause[2,3], Mark A. Brzezinski[4], Bethanie R. Edwards[5], William P. Biggs[1], Michael Maniscalco[4], John P. McCrow[6], Benjamin A.S. Van Mooy[7], Kay D. Bidle[1], Andrew E. Allen[6,8], Kimberlee Thamatrakoln[1*]

[1] Department of Marine and Coastal Sciences, Rutgers University, New Brunswick NJ, USA.
[2] Dauphin Island Sea Lab, Dauphin Island, AL, USA.
[3] Department of Marine Sciences, University of South Alabama, Mobile, AL, USA.
[4] Marine Science Institute, University of California, Santa Barbara, CA, USA.
[5] Department of Earth and Planetary Science, University of California, Berkeley, CA, USA.
[6] Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA, USA.
[7] Woods Hole Oceanographic Institution, Woods Hole, MA, USA.
[8] Scripps Institution of Oceanography, University of California, San Diego, CA.

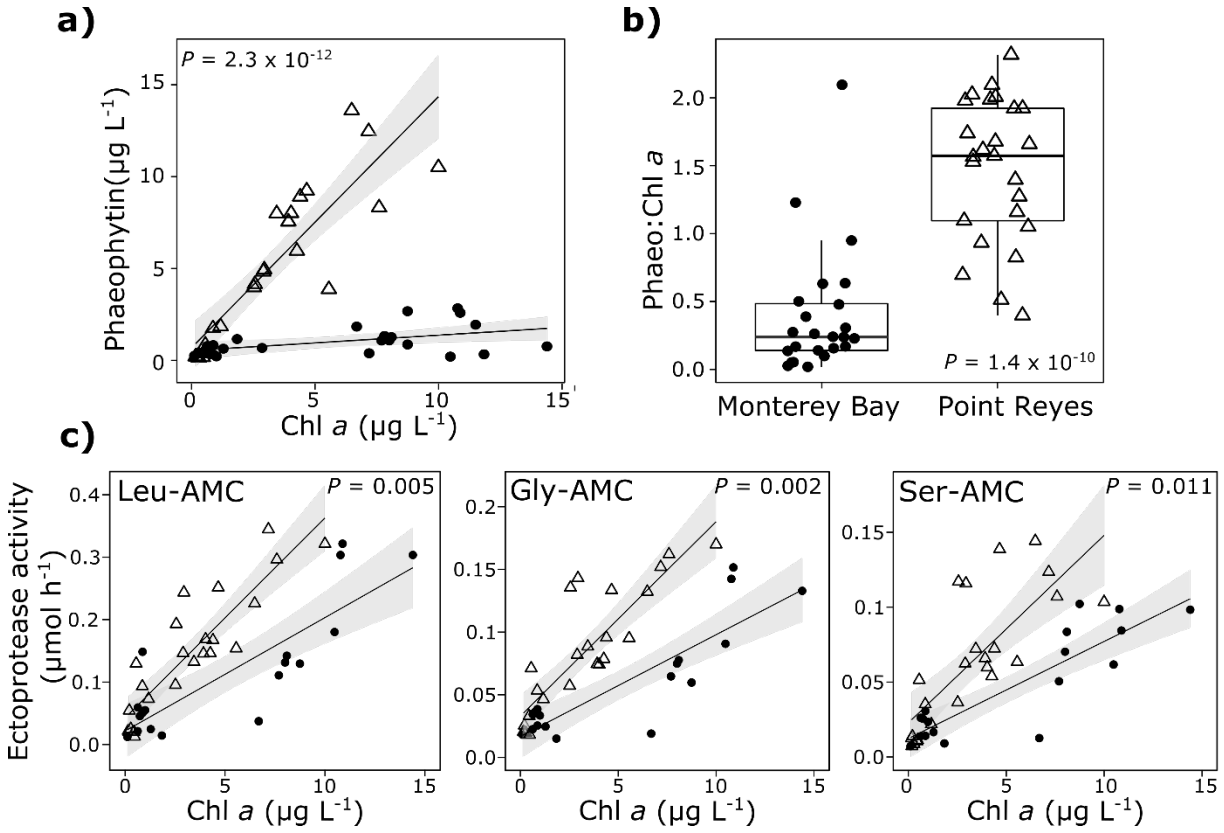
*Correspondence to: thamat@marine.rutgers.edu

This PDF file includes:

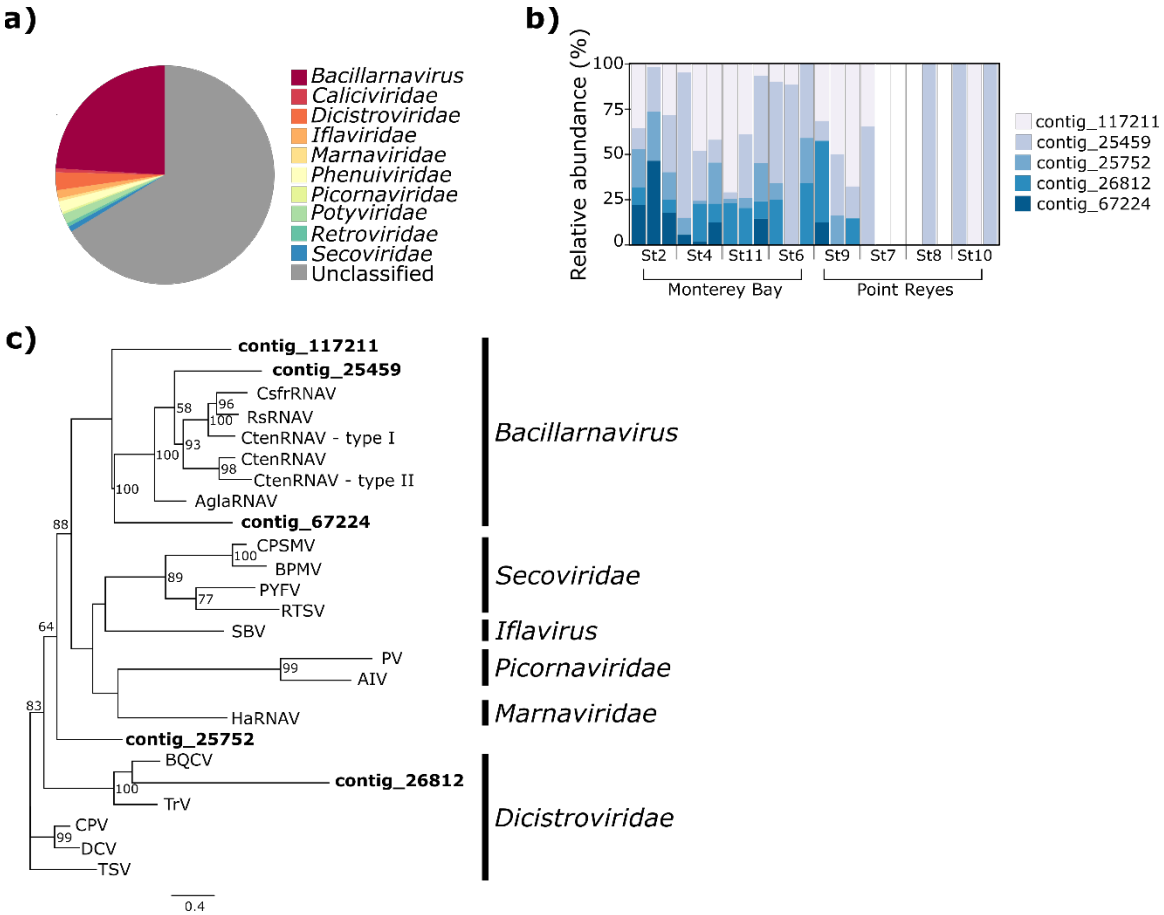
Supplementary Figures 1 to 7
Supplementary Tables 1 to 3

Other Supplementary Materials for this manuscript include:

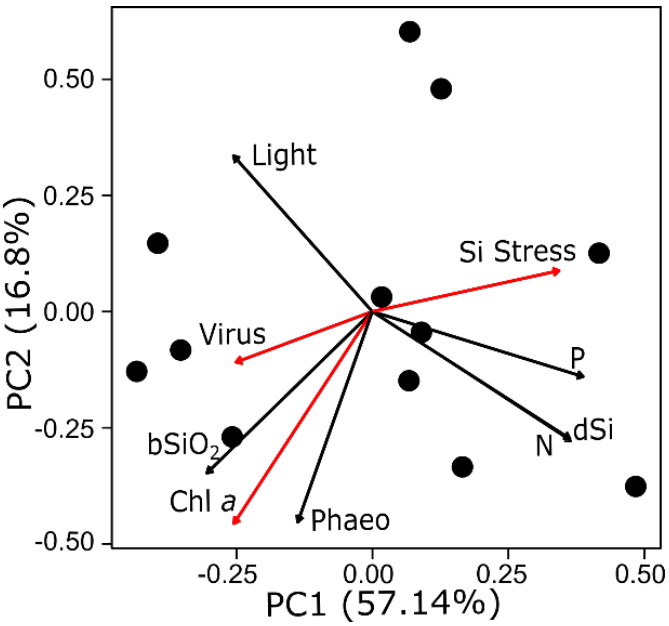
Supplementary Data 1: Metatranscriptome and diatom virus analysis



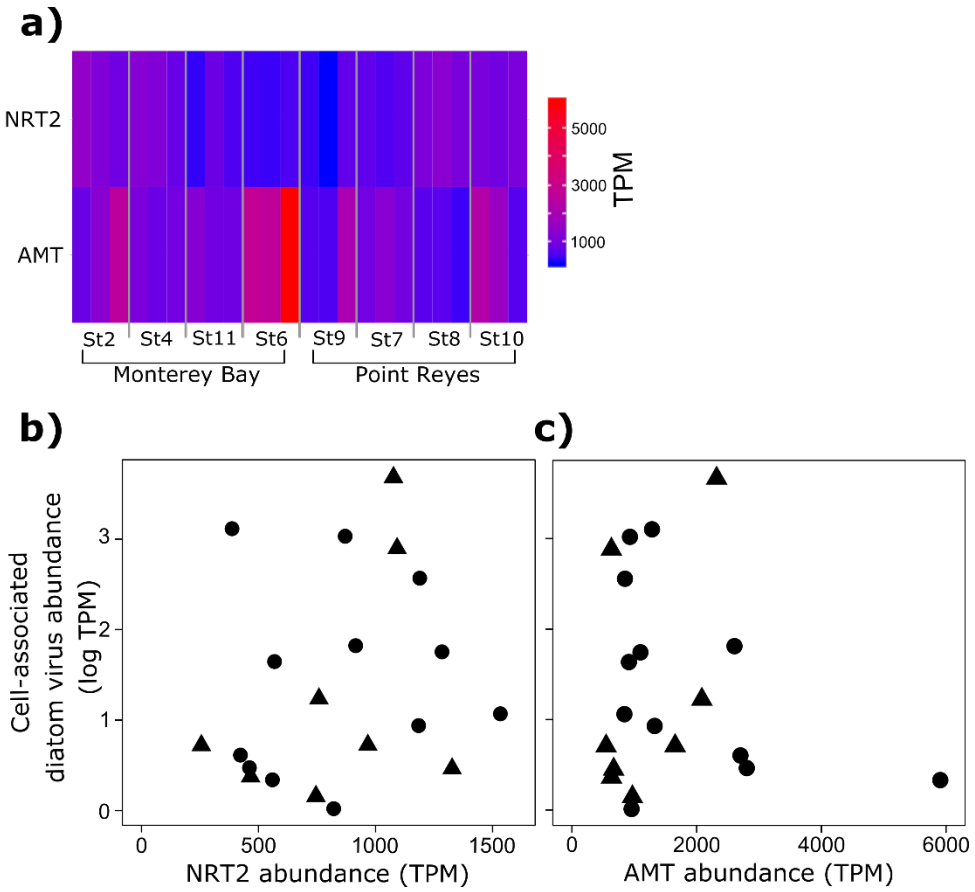
Supplementary Figure 1. Signatures of mortality in Point Reyes. a, The relationship between phaeophytin and chl *a* in Monterey Bay (closed circles) and Point Reyes (open triangles). Lines of best fit with 95% confidence bands (grey shading) depict linear regression of the data with r^2 = 0.78, $P = 5.7 \times 10^{-9}$ for Point Reyes (analysis of variance; n = 25 independent samples); r^2 = 0.197, P = 0.02 for Monterey Bay (analysis of variance; n = 24 independent samples). **b,** Box and whisker plot of phaeophytin:chl *a* in Monterey Bay and Point Reyes (n = 24 and 25 independent samples, respectively), depicting the median (horizontal line), upper and lower quartiles of the data, with whiskers encompassing data points within 1.5x of the interquartile range. P value was determined using Welch's two sample, two-sided t-test. **c,** The relationship between ectoproteolytic activity (µmol h^{-1}) and chl *a* for Monterey Bay (closed circles; n = 13 independent samples) and Point Reyes (open triangles; n = 18 independent samples), assayed via hydrolysis of leucine-AMC, glycine-AMC, and serine-AMC. Lines of best fit with 95% confidence bands (grey shading) depict linear regression of the data with r^2 = 0.79 and 0.70 (Leu-AMC); r^2 = 0.75 and 0.75 (Gly-AMC); r^2 = 0.59 and 0.77 (Ser-AMC) for Point Reyes and Monterey Bay, respectively (with all P values < 0.001). P values in panels a and c describe the statistical significance between the linear regressions of the data from Monterey Bay and Point Reyes (analysis of variance).



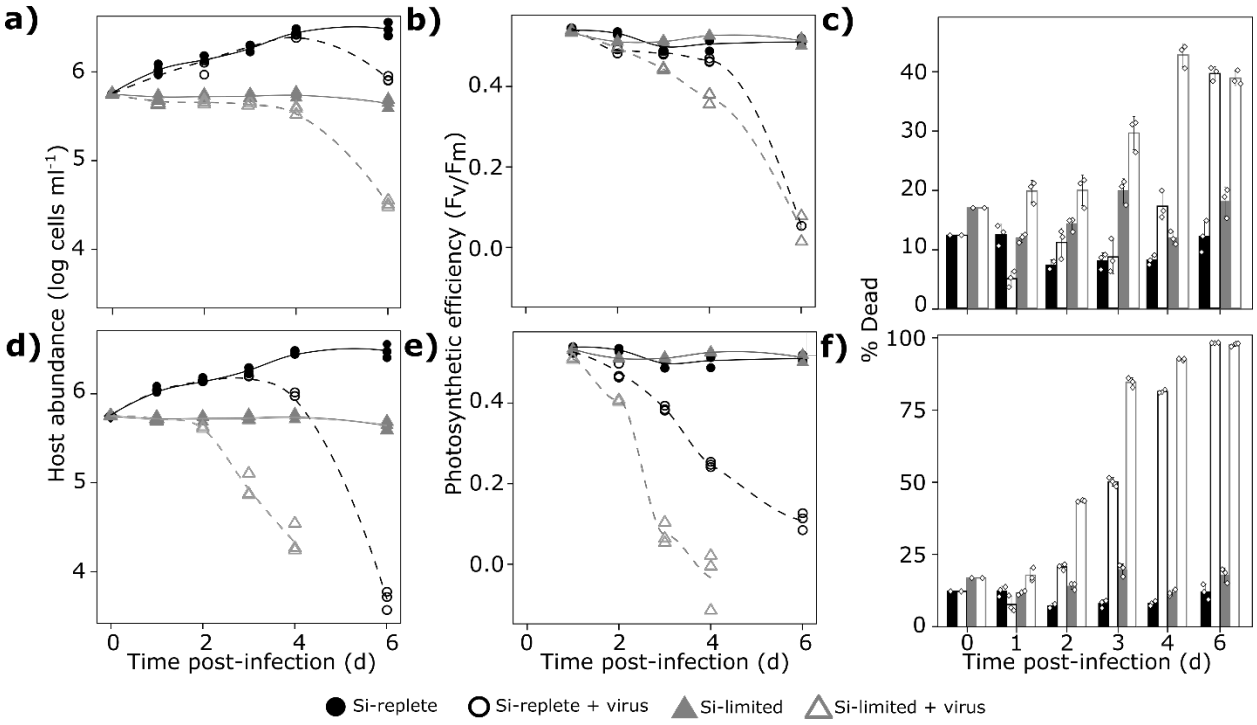
Supplementary Figure 2. Abundance and identification of diatom ssRNA virus contigs. a, Relative abundance of ssRNA viruses, grouped by family, identified within the metatranscriptome across all stations and depths. **b,** Relative abundance (%) of five cell-associated diatom virus-like contigs at three depths corresponding to 55, 22 and 1% surface irradiance, shown within each station from left to right. Open, white bars represent stations in which no diatom virus-like contigs were detected. **c,** Maximum likelihood phylogenetic tree of RdRP amino acid sequences within the order *Picornavirales* and homologous contigs identified within the metatranscriptome. Bootstrap values > 50 are shown (100 replicates). The abbreviations, names and NCBI database accession numbers of the amino acid sequences that were used to construct the reference alignment are: CsfrRNAV, *Chaetoceros socialis radians* RNA virus 1, YP_002647032.1; RsRNAV, *Rhizosolenia setigera* RNA virus 01, YP_006732323.1; CtenRNAV – type I, *C. tenuissimus* RNA virus 01, YP_009505620.1; CtenRNAV-type II, *C. tenuissimus* RNA virus type-II, BAP99818.1; AglaRNAV, *Asterionellopsis glacialis* RNA virus, BAP16719.1; CPSMV, cowpea severe mosaic virus, NP_619518.1; BPMV, Bean pod mottle virus, NP_612349.1; PYFV, Parsnip yellow fleck virus, BAA03151.1; RTSV, Rice tungro spherical virus, NP_042507.1;SBV, Sacbrood virus, AIZ75645.1; PV, Human poliovirus 1, CAA24461.1; AIV, Aichi virus 1, ADN52312.2; HaRNAV, *Heterosigma akashiwo* RNA virus, AAP97137.1; BQCV, Black queen cell virus, NP_620564.1; TrV, Triatoma virus, NP_620562.1; CPV, Cricket paralysis virus, NP_647481.1; DCV, *Drosophila* C virus, NP_044945.1; TSV, Taura syndrome virus, NP_149057.1.



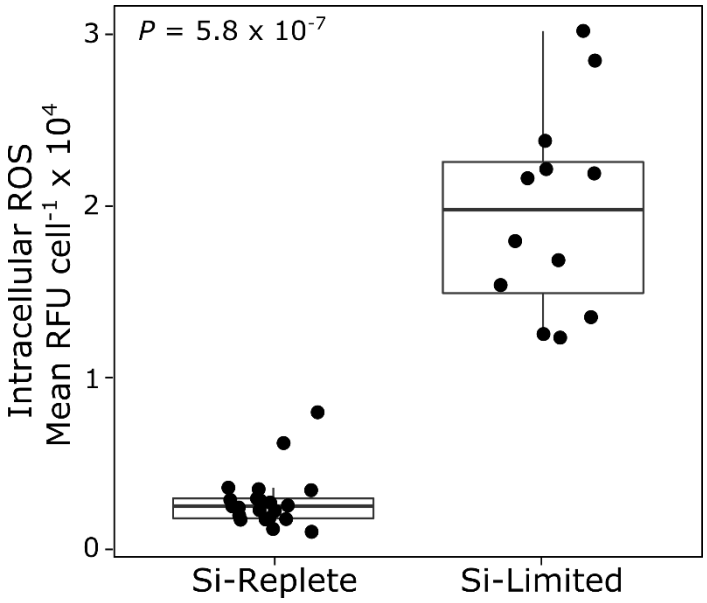
Supplementary Figure 3. **Principal component analysis of Monterey Bay variables.** Chemical and biological data (n = 12 independent samples) include dissolved nutrients (N, P, dSi), phaeophytin (Phaeo), chlorophyll *a* (Chl *a*), biogenic silica ($bSiO_2$), Si stress, % surface irradiance (light) and normalized cell-associated diatom virus abundance (Virus). Higher Si stress values are indicative of Si-replete conditions. Red vectors indicate significant correlation ($P < 0.05$, analysis of variance) using linear regression analysis of each variable with normalized cell-associated diatom virus abundance. Precise r^2, slopes and *P* values for each linear regression are summarized in Supplementary Table 2.



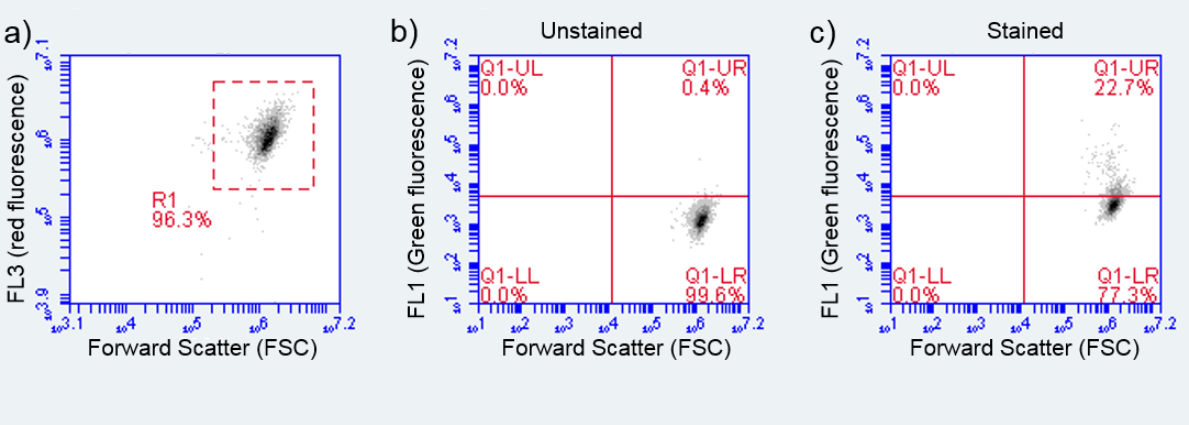
Supplementary Figure 4. Assessing nitrate limitation in the California Current Ecosystem. **a,** Relative abundance of diatom-specific NRT2 (nitrate transporter) and AMT (ammonium transporter) transcripts at each station. Total reads were group-normalized and are expressed as transcripts per million (TPM). Data from three depths, corresponding to 55, 22 and 1% surface irradiance are shown from left to right within each station. **b-c,** Normalized cell-associated diatom virus abundance (log TPM) vs. NRT2 **(b)** and AMT **(c)** transcript abundance (TPM). Circles and triangles represent Monterey Bay and Point Reyes stations, respectively.



Supplementary Figure 5. The impact of Si limitation on diatom host-virus dynamics in laboratory cultures of *Chaetoceros tenuissimus*. a,d, Host abundance and **b,e,** photosynthetic efficiency (F_v/F_m) in Si-replete (black lines and circles) and Si-limited cultures (grey lines and triangles) infected with CtenRNAV (**a-c**) and CtenDNAV (**d-f**). Control, uninfected cultures are depicted by closed symbols; infected cultures are open symbols. Each biological replicate (n=3) is shown by individual symbols, with lines of best fit depicting a LOESS regression. **c,f,** Percent dead cells determined using SYTOX staining and flow cytometry. Black bars are Si-replete cultures and grey bars are Si-limited cultures (closed, control; open, infected). Individual data points and the mean ± standard deviation are shown. Data are representative of one to four biologically independent experiments.



Supplementary Figure 6. Reactive oxygen species production during Si limitation. Intracellular reactive oxygen species (ROS) in biologically independent cultures of Si-replete (n = 21) and Si-limited (n = 12) *Chaetoceros tenuissimus*, assessed via H_2DCFDA staining and flow cytometry. Box plots depict the median (horizontal line), upper and lower quartiles of the data, and whiskers encompassing data points within 1.5x of the interquartile range. *P* value was determined using Welch's two sample, two-sided t-test.



Supplementary Figure 7. **Gating strategy of flow cytometry data.** a, Cell abundance was measured using flow cytometry (BD Accuri C6;) at a flow rate of 14 µL min^{-1} with a 10 µm core diameter. Cell populations were identified by plotting chlorophyll fluorescence (488 nm excitation, >670 nm emission) and forward scatter. **b,c,** For stained samples, preliminary gating of the population was determined by forward scatter (dashed box shown in panel a) and green fluorescence was measured (488 nm excitation, 533/30 emission). Unstained samples were used to set a threshold for background green fluorescence (b). The number of positively stained cells in a population was determined as those that exceeded the background fluorescence of the unstained sample (c).

Station	Region	%I_0	PO_4	$NO_3 + NO_2$	dSi	Chl *a*	Phaeo	$bSiO_2$	ρ_{amb}	ρ_{enh}	V_{amb}	V_{enh}
2	MB	55	0.41	8.79	9.98	8.74	0.88	5.28	3.76	3.98	0.62	0.65
2	MB	22	0.43	8.11	10.8	8.09	1.28	5.52	3.50	3.70	0.56	0.59
2	MB	1	0.64	11.1	14.2	7.69	1.10	5.63	0.54	0.71	0.09	0.12
4	MB	55	0.09	0.23	1.60	10.9	2.60	6.69	2.52	11.1	0.35	1.22
4	MB	22	0.11	0.63	2.11	14.4	0.78	5.72	3.03	10.7	0.47	1.34
4	MB	1	0.38	7.65	7.56	10.5	0.22	4.57	0.98	0.79	0.20	0.17
11	MB	55	0.07	0.59	2.85	11.8	0.35	5.56	4.30	8.52	0.66	1.16
11	MB	22	0.11	3.23	5.73	8.74	2.69	5.13	5.64	8.55	0.89	1.24
11	MB	1	0.94	18.3	20.2	7.81	1.34	0.85	0.81	0.75	0.78	0.73
6	MB	55	0.36	3.86	6.49	0.75	0.36	1.24	0.71	0.83	0.51	0.58
6	MB	22	0.38	4.46	6.32	1.02	0.23	1.14	0.69	0.82	0.53	0.62
6	MB	1	0.78	11.7	13.8	1.30	0.65	0.27	0.09	0.06	0.32	0.21
9	PR	55	0.22	0.13	3.54	0.18	0.22	0.21	0.08	0.10	0.36	0.45
9	PR	22	0.20	0.08	3.46	0.20	0.23	0.21	0.07	0.10	0.31	0.43
9	PR	1	0.66	9.07	12.6	0.49	0.40	0.31	0.01	0.01	0.04	0.04
7	PR	55	0.12	1.57	1.64	3.93	7.54	7.63	1.61	15.2	0.19	1.15
7	PR	22	0.11	1.07	1.51	4.03	8.01	8.73	1.36	8.58	0.15	0.71
7	PR	1	0.27	4.29	3.44	4.27	5.96	8.45	1.21	1.22	0.13	0.14
8	PR	55	0.09	0.04	0.47	7.17	12.5	16.9	0.07	14.5	0.00	0.72
8	PR	22	0.10	0.10	0.80	9.99	10.5	17.1	0.62	3.70	0.04	0.21
8	PR	1	0.75	10.8	17.9	5.56	3.88	7.37	0.25	0.52	0.03	0.07
10	PR	55	0.08	0.01	0.48	2.57	4.16	n.d.	0.34	2.94	n.d.	n.d.
10	PR	22	0.05	0.05	0.52	2.95	4.95	4.61	0.33	4.58	0.07	0.81
10	PR	1	0.30	1.96	3.24	6.49	13.6	21.7	2.63	6.77	0.12	0.29

Supplementary Table 1. DYEatom cruise metadata. Chemical and biological parameters measured at each station in Monterey Bay (MB) and Point Reyes (PR) during the DYEatom cruise. Station number, region, and % surface irradiance (%I_o) are shown with dissolved nutrients (µmol L^{-1}; phosphate, PO_4; nitrate and nitrite, $NO_3 + NO_2$; silicon, dSi), chlorophyll *a* (Chl *a*, µg L^{-1}), phaeophytin (Phaeo, µg L^{-1}), biogenic Si ($bSiO_2$, µmol L^{-1}),^{32}Si uptake rates at ambient [$Si(OH)_4$] (ρ_{amb}) and with a 18 µM enhancement (ρ_{enh}, µmol Si L^{-1} day^{-1}) and specific Si uptake rates (ρ * $bSiO_2^{-1}$) at ambient [$Si(OH)_4$] (V_{amb}) and with a 18 µM enhancement (V_{enh}, day^{-1}); n.d. indicates no data.

Parameter	Monterey Bay			Point Reyes		
	r^2	*P*	slope	r^2	*P*	slope
% I_0	0.053	0.473	0.008	0.010	0.816	-0.005
dSi	0.107	0.299	-0.050	0.003	0.891	-0.015
$N0_3$+NO_2	0.149	0.215	-0.060	0.001	0.951	0.009
PO_4	0.245	0.102	-1.503	0.002	0.918	0.234
Phaeophytin	0.222	0.122	0.477	0.066	0.540	0.054
$bSiO_2$	0.278	0.078	0.192	0.313	0.192	0.048
Chl a	**0.389**	**0.030**	**0.120**	0.005	0.874	0.021
Si Stress	**0.430**	**0.021**	**-2.218**	0.038	0.644	-0.592

Supplementary Table 2. Linear regression analysis. Summary of linear regression analysis between chemical and biological metadata and normalized cell-associated diatom virus abundance in Monterey Bay and Point Reyes. Metadata (found in Table S1, n = 24 independent samples) include surface irradiance (%I_0), dissolved silicon (dSi, µmol L^{-1}), dissolved nitrate and nitrite (NO_3 + NO_2, µmol L^{-1}), dissolved phosphate (PO_4, µmol L^{-1}), phaeophytin (µg L^{-1}), biogenic silica ($bSiO_2$, µmol L^{-1}), chlorophyll *a* (Chl *a*, µg L^{-1}) and Si Stress (V_{amb}:V_{enh}). Bold values are statistically significant relationships ($P < 0.05$, analysis of variance) with normalized cell-associated diatom virus abundance.

Contig	Primer Sequence (5'-3')	$T_{annealing}$ (°C)
25459	TGCGCTAGTGCAGCAAGA[a]	54
	AGGCATCGCAACAACCAC[b]	54
	GTTCACCTAATTCTGCCCATACTG[b]	54
	TCAGTAGCACCATTGGCGTG[a]	54
67224	CTGGCTAAGAAGGGTCTCAG[a]	53
	AGCAATCTCGCTGGCAATAG[b]	52
	GGAGCTGGATCAAACACCTATCC[a,b]	53[a], 52[b]
117211	GCGTCCGATGCCTCGCTAG[a]	55
	GATGCCGATTGGTTGCCG[b]	54
	AACGCGCAAGTTGGCC[a,b]	55[a], 54[b]

[a]outer primer pair
[b]nested or semi-nested primer pair

Supplementary Table 3. Primers used for extracellular diatom virus quantification. Primer sequences and annealing temperatures for targeted quantitative reverse-transcription PCR of each diatom virus contig identified in the cell-associated fraction via metatranscriptomic analysis. Primer pair amplification efficiencies, determined by qPCR with a dilution series of purified product (1-1000 copies), ranged between 97.4 – 103.2%.